Context-Aware Text Representation for Social Relation Aided Sentiment Analysis

Minh Luan Nguyen
Institute for Infocomm Research
Singapore
mlnguyen@i2r.a-star.edu.sg

ABSTRACT

In this paper, we propose CaTER, which learns a novel context-aware joint representation of text and user by incorporating semantic text embedding of unlabeled tweets as well as social relation information. CaTER leverages the wealth of user contextual information available apart from user's utterances for sentiment analysis. Our approach is inspired by social science about emotional behaviors of connected users, who perhaps more likely to consensus on similar opinions. Our method outperforms numerous baselines on two real-world Twitter datasets.

Keywords

sentiment analysis, deep learning, text embedding, social relations

1. INTRODUCTION

Sentiment analysis for social media platforms like Twitter poses several challenges for researchers. The social text data often contains short and noisy messages. It also contains lots of emotional abbreviations, emoticons and has no syntactic structure. Because of highly discourse variations, sentiment data in social media often lacks of sufficient labeled data. However, existing sentiment analysis methods require sufficient texts labeled with polarity, thereby, not suitable for social media data. Currently, there are some effort to overcome this issue. Hu et al. [2] used supervised method to model message similarity by exploiting user relationship. Go [1] used distant supervision based on noisy labels. Speriosu et al. [4] used semi-supervised label propagation on unlabeled tweets upon social relations. Tang [5] used deep learning to learn an effective word representation based on noisy labelled tweets.

Besides textual content, networked characteristic is a distinct feature of blog messages via user relationships because it may contain useful semantic information, which are not available in traditional purely text contents. According to the principle of homophily [2], when user are connected by a social relationship, they might tend to consensus on opinions. Motivated by this, we propose a three-phase semi-supervised deep learning framework, called CaTER, for sentiment analysis in social media. CaTER leverages on both ground-truth labeled, unlabeled tweets and social relation data in tackling the social media sentiment issues. CaTER is a semi-supervised method, which works on both of text-level and user-level. We use a semi-supervised approach in order to exploit the vast amount of unlabeled data to capture semantic similarity of textual and user-context via latent low-dimensional space representations.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518/2889347.

2. PROPOSED APPROACH

Given a corpus $D = \{t_1, t_2, ..., t_N\}$ of N tweets, which are generated by the set of users $U = \{u_1, u_2, ..., u_M\}$, where M is the number of distinct users. Let $W = \{w_1, v_2, ..., w_K\}$ be the word vocabulary. For each message in the corpus $t_i \in D$, $t_i = (x_i, y_i) \in \mathbb{R}^{n+c}$ consists of tweet and sentiment label, where $x_i \in \mathbb{R}^n$ is the tweet feature vector and $y_i \in \mathbb{R}^c$ is the sentiment label vector. Learning a sentiment classifier can then be abstracted into two main steps. Firstly, we find a feature embedding function Φ that maps a tweet t posted by user u to its feature vector $e_{x,u} = \Phi(x,u)$. Secondly, we learn a sentiment classifier $f(\Phi(x,u))$ to automatically assign sentiment labels for unseen message t posted by user u (i.e., test data) such that $f(\Phi(x,u)) = y$. This paper focuses on learning a good user-context-aware text representation $\Phi(x,u)$.

We propose a 3-phase semi-supervised deep learning framework, called Context-aware Text Representation (CaTER), for sentiment analysis in social media.

Text Representation. We learn an embedding function $\Phi_T(x) : \mathbb{R}^n \to \mathbb{R}^d$ that represents the semantic meaning of each tweet in a vector space. Each word $w \in W$ is associated with a d-dimensional vector, i.e., $e_w \in \mathbb{R}^d$. We use the *skip-gram* model proposed in the word2vec[1] algorithm to learn the word presentation. In this model, we are given an unlabelled tweet corpus T of words $w \in W$ and their contexts c of neighboring words. Let $p(c|w)$ be the conditional probabilities of the context of a given word w. To learn word representation given the corpus T, the goal is to set the parameters $\Phi_T(w) = e_w$ of $p(c|w; \Phi_T)$ to maximize the corpus probability

$$\underset{\Phi_T}{\text{maximize}} \prod_{w \in T} [\prod_{c \in C(w)} p(c|w; \Phi_T)]$$

where $C(w)$ is the set of contexts of word w. The above objective function is maximized using Negative Sampling technique. Finally, the vector representation e_x of a tweet sequence $x = (w_1, ..., w_k)$ is computed based on its word's embedding via a nonlinear mapping.

User-Context Representation. It learns an embedding function $\Phi_U(u) : \mathbb{R}^M \to \mathbb{R}^d$ that represents the semantic meaning of each user-context from social relation graph. Each user $u \in U$ is associated with a d-dimensional vector, i.e., $e_u \in \mathbb{R}^d$.

Let $G = (U, E)$ be the social relation graph, where U is the user list and E is the connections. For sentiment homophily, we want to learn a community-aware latent representation, in which the distance between latent embedding of users should preserve the latent similarity between the corresponding users in the social graph. We

[1] https://code.google.com/p/word2vec/

learn user representation in a similar way for text representation by using random walks. Let $W_{u_i} = \{u_{i-p}, .., u_i, .., u_{i+p}\}$ be the random walk centered at user u_i. Random walk is widely used to model the node similarity for community detection because it captures statistical structure information of network. We model the graph as a collection of fixed-length random walks. These walks can be considered equivalently as short sentences and phrases in a text corpora. Hence, similar as in text, we can estimate the likelihood of observing the context $c(u_i)$ of a given user u_i in the random walk. It maximizes the objective probability of any user appearing in the random walk context using stochastic gradient descent:

$$\underset{\Phi_U}{\text{maximize}} \prod_{u \in G} \log \Pr(\{u_{i-p}, \ldots, u_{i+p}\} \setminus u_i | \Phi_U(u_i))$$

Context-aware Text Representation. We propose a bilinear embedding model that learns a joint d-dimensional embedding space for context-aware user-text representation using labeled data. Let $e_x = \Phi_T(x), x \in \mathbb{R}^n$ be a text embedding and let $e_u = \Phi_U(u), u \in \mathbb{R}^M$ be a user embedding. We have useful information about user embedding responsible for each tweet post. The joint user-text model leverages text representation by incorporating a user dependent bias term into. This assumption captures the notion of homophily: users represented by similar vectors are more likely to write similar text. As user information (i.e., interest, background, etc...) is reflected in their user generated text (e.g., a sports fan is more likely to talk sports), predicting user-generated content given users' representations impacts the learning of text representations. In particular, the context-aware user-text joint embedding, $e_{x,u} = \Phi_{T+U}(x,u) : \mathbb{R}^n \times \mathbb{R}^M \rightarrow \mathbb{R}^d$, is defined as

$$\Phi_{T+U}(x,u) = Pe_x + Qe_u$$

where $P \in \mathbb{R}^{d \times n}$ and $Q \in \mathbb{R}^{d \times M}$ are text and the user embedding matrix respectively. We regularize the model by constraining the norms of P and Q to prevent the model from overfitting: $\|P_i\|_2 \leq, i = 1, \ldots, n$, and $\|Q_i\|_2 \leq 1, i = 1, \ldots, M$. Recall that $y \in Y = \{1, \ldots, c\}$ denote the list of sentiment labels. Let $\Phi_Y(y) : \mathbb{R}^c \rightarrow \mathbb{R}^d$ be label embedding. The bilinear model is of the form:

$$g([x,u],y) = \Phi_{T+U}(x,u)^T \Phi_Y(y)$$

Training algorithm. We train our bilinear embedding model $g([x,u],y)$ to learn Φ_{T+U} and Φ_Y by minimizing the Weighted Approximate Rank Pairwise loss defined as in [6]. This loss approximately optimizes the loss function using stochastic gradient descent with a negative sampling technique.

3. EXPERIMENTAL EVALUATION

We evaluate our algorithm on two corpora: Stanford Twitter Sentiment (STS) and Obama-McCain Debate (OMD). The Stanford Twitter Sentiment (STS) was collected by Go et al. [1], which consists of 8467 users and 40216 tweets with polarity sentiment labels. The Obama-McCain Debate (OMD) consists of 735 users and 3,269 tweets posted during the presidential debate on September 26, 2008 [4]. Due to original purpose, both of the datasets lack social relation graph information between users in this dataset. And the number of unlabelled tweets is also not sufficient for effective learning of word representation using deep learning. To overcome this, we exploit a complementary and very-large Twitter social network dataset[2] [3], which consists of 11 millions of users, 1.46 bilions of social relations and 106 millions of tweets. We use the Twitter complete follower graph to obtain the social relation graph. In addition, the initial word representation is learnt using this unlabelled tweets. We compare our framework with several state-of-the-art *tf-idf* baselines for sentiment analysis including SVM, Distant Supervision (DS), Label Propagation (LPROP) and SANT [2] and one Logistic Regression with word embedding (LG-we) [5]. For CaTER, we train our model with simple linear logistic regression $f(e_{x,u})$ based on the final text embedding $e_{x,u} = \Phi_{T+U}(x,u)$.

[2]http://an.kaist.ac.kr/traces/WWW2010.html

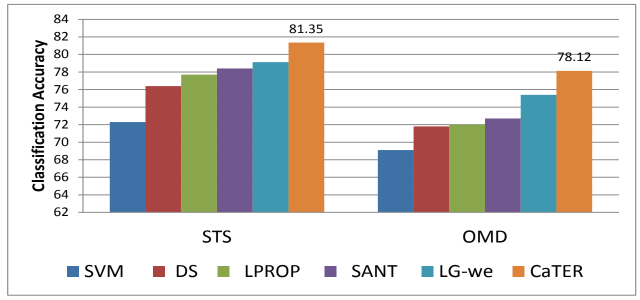

Figure 1: Sentiment accuracy comparison on STS and OMD.

Figure 1 presents the sentiment classification accuracy results on STS and OMD datasets. We find that the proposed method CaTER has the best accuracy among all the other methods. We also observe that although the noisy labels are helpful on sentiment task, using only noisy labels to train a classifier (DS) marginally increases the accuracy as compared with supervised ground-truth methods like SVM. We also see that LPROP and SANT, which exploit both text and social graph information, can also only gain a considerable accuracy improvement as compared with traditional SVM and DS. More importantly, we find that LG-w2v, which uses only low-dimensional latent representation for text, can also only gain a marginal accuracy improvement as compared with LPROP and SANT. This emphasizes the effectiveness of our Phase 1 for text embedding. In particular, our CaTER has achieved a superior performance on both datasets. Specifically, CaTER outperforms the best SANT baseline by 3.0% and a text embedding by 5.2% on average. It shows that incorporating user-context information into text representation helps improve the performance.

4. CONCLUSIONS

Our experiments establish the effectiveness of exploiting user-context information for leveraging social media sentiment analysis using social relation graph. When labeled data is scarce, vast amount of unlabelled data and social relation graph are available, deep learning and latent low-dimensional vector representation can be an effective way to do semi-supervised sentiment analysis.

5. REFERENCES

[1] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12–, 2009.

[2] X. Hu, L. Tang, J. Tang, and H. Liu. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of WSDM*, pages 537–546. ACM, 2013.

[3] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th conference on WWW*, pages 591–600. ACM, 2010.

[4] M. Speriosu and J. Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Workshop on Unsupervised Learning in NLP*, pages 53–63. ACL, 2011.

[5] D. Tang, M. Zhou, and B. Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the ACL*, pages 1555–1565, 2014.

[6] J. Weston, S. Bengio, and N. Usunier. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, volume 11, pages 2764–2770, 2011.